Journal of the Formosan Medical Association (2013) 112, 332–337

Available online at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.jfma-online.com

ORIGINAL ARTICLE

The test–retest reliability and the minimal detectable change of the Purdue pegboard test in schizophrenia

Posen Lee [a], Chin-Hsuan Liu [b], Chia-Wei Fan [c], Chi-Pang Lu [c], Wen-Shian Lu [d,e,*], Ching-Lin Hsieh [c,f]

[a] *Department of Occupational Therapy, I-Shou University, Kaohsiung, Taiwan*
[b] *Department of Occupational Rehabilitation, Kai-Suan Psychiatric Hospital, Kaohsiung, Taiwan*
[c] *School of Occupational Therapy, College of Medicine, National Taiwan University, Taipei, Taiwan*
[d] *School of Occupational Therapy, Chung Shan Medical University, Taichung, Taiwan*
[e] *Occupational Therapy Room, Chung Shan Medical University Hospital, Taichung, Taiwan*
[f] *Department of Physical Medicine and Rehabilitation, National Taiwan University Hospital, Taipei, Taiwan*

Received 9 September 2010; received in revised form 16 February 2012; accepted 17 February 2012

KEYWORDS
hand;
psychometrics;
Purdue pegboard test;
reproducibility of results;
schizophrenia

Background/Purpose: The Purdue pegboard test is widely used in measuring the hand dexterity of patients with schizophrenia. In patients with schizophrenia, the test–retest reliability and minimal detectable change (MDC) of this test remain largely unknown, limiting the interpretability of this popular measure. The purpose of this study was to estimate the test–retest reliability and the MDC of the Purdue pegboard test for patients with schizophrenia.

Methods: A total of 147 patients with schizophrenia participated in this study. The participants were administrated the five subtests of the Purdue pegboard test, three trials in a row at both of the two sessions 1 week apart. The intraclass correlation coefficient (ICC) was used to examine the test–retest reliability and the MDC was calculated on the basis of standard error of measurement.

Results: The test–retest reliabilities of the five subtests were moderate to good (ICC = 0.73–0.88). The MDC (MDC%) was 3.0 (22.9%) for the dominant hand subtest, 3.1 (26.1%) for the nondominant hand subtest, 3.0 (31.7%) for the both hands subtest, 6.1 (17.7%) for the dominant + nondominant + both hands subtest, and 8.5 (35.3%) for the assembly subtest.

Conclusion: Our results reveal that the Purdue pegboard test has moderate-to-good test–retest reliability but substantial random measurement error. These findings should

* Corresponding author. School of Occupational Therapy, Chung Shan Medical University, No. 110, Section 1, Jianguo North Road, Taichung City 40201, Taiwan.
E-mail address: wslu@csmu.edu.tw (W.-S. Lu).

doi:10.1016/j.jfma.2012.02.023

enable clinicians and researchers to monitor and interpret the changes in the hand dexterity of patients with schizophrenia more accurately and confidently.

Introduction

By doing everyday tasks with his/her hands, a person can gain competence and master his/her desired and expected roles in life.[1] Hand dexterity deficits are common in patients with schizophrenia.[2] Dexterity deficits often limit or restrict the individual's capacity to complete daily life tasks effectively and efficiently.[3] In the investigation of the effects of treatments for patients with schizophrenia, hand dexterity is usually an important indicator because of its close relevance to executive function and impact on social function.[4] Hand dexterity also influences employment outcomes for patients with schizophrenia.[5] Clinicians assess the hand dexterity of patients with schizophrenia periodically to revise treatment programs in order to improve the patients' ability to perform in their occupations. Thus, a reliable instrument is crucial in measuring hand dexterity function in patients with schizophrenia to monitor the effects of intervention in both clinical and research settings.

To examine the reliability of a measure, the test–retest reliability can be calculated by the extent of agreement and reproducibility between two repeated measurements.[6] The minimal detectable change (MDC) is defined as the minimal threshold beyond the random measurement error with a 95% confidence level.[7] Namely, MDC (absolute value) indicates the minimal magnitude of change beyond which the change is likely to be real, rather than due to random measurement error. Both clinicians and researchers can use the MDC as a threshold to judge whether a change in an individual patient on a certain measure referring to a specific characteristic signifies real improvement (or deterioration). Thus, the MDC is critical in interpreting individual changes in successive measurements in both clinical and research settings.

The test–retest reliability of the Purdue pegboard test has been examined in healthy persons and persons diagnosed with multiple sclerosis, rheumatoid arthritis, and mental retardation.[8–12] However, the test–retest reliability and the MDC of the Purdue pegboard test when administered to patients with schizophrenia have not been examined and thus remain largely unknown. Because psychometric properties (e.g., test–retest reliability) are sample dependent,[13,14] such a shortcoming of the Purdue pegboard test limits the measure's utility and interpretability for patients with schizophrenia.

The purpose of this study was to examine the test–retest reliability and the MDC of the five subtests of the Purdue pegboard test in patients with schizophrenia.

Methods

Participants

Participants were recruited from a chronic ward of a clinical psychiatric center in southern Taiwan. The eligible participants met the following criteria: (1) diagnosis of schizophrenia according to the Diagnostic and Statistical Manual of Mental Disorders, 4th ed. (DSM-IV) criteria, excluding schizoaffective disorder; (2) stable and consistent dose of antipsychotic medication for at least 1 month prior to the study; (3) ability to follow instructions; and (4) absence of substance abuse or other neurological deficits such as dementia, mental retardation, or developmental disability. All participants gave informed consent prior to their inclusion in this study and were assigned identification numbers to maintain anonymity. Data were collected from February 2009 to May 2009. The study protocol was reviewed and approved by the Institutional Review Board of Kai-Suan Psychiatric Hospital.

Procedures

The Purdue pegboard test was administered at two sessions 1 week apart by a trained occupational therapist. All participants completed three trials per session. The therapist followed the standard procedure and gave the instructions according to the Purdue pegboard test manual. The test was administered by group, and each group comprised no more than five patients. Participants' demographic characteristics were collected from medical charts.

Instrument

The Purdue pegboard test, which has good predictive and concurrent validity,[15,16] consists of five subtests: dominant (D) hand, nondominant (ND) hand, both (B) hands, dominant + nondominant + both (D + ND + B) hands, and assembly subtest. On a Purdue pegboard, each of the extreme right- and left-side cups contains 25 pins. For right-handed participants, the cup next to the extreme right cup contains 20 collars, and the cup next to the extreme left cup contains 40 washers; this arrangement is reversed for left-handed participants. The test can be administrated either individually or by group. In the D hand and ND hand subtests, participants are instructed to place as many pins as possible in the holes in 30 seconds. In the B hands subtest, participants use the D and ND hands simultaneously to place the pins in both holes in 30 seconds. In the assembly subtest, participants pick up and place pins, washers, and collars, using alternative hands, for 60 seconds. The number of pins placed in the pegboard within the time limit represents the scores for the D hand and ND hand subtests, and the pairs of pins for the B hands subtest. The sum of scores of the three subtests represents the score for the D + ND + B hands subtest. The number of pieces assembled completely (containing pin, washer, collar, and second washer) represents the score of the assembly subtest.[17]

Statistical analysis

The mean scores of the three trials in each session of the five subtests were computed for the test and retest

sessions. The one-way analysis of variance (ANOVA-based intraclass correlation coefficient ($ICC_{2,1}$) was used to analyze the test–retest reliability of the Purdue pegboard test.[6] The reason for using the ICC rather than Pearson correlation is that the Pearson correlation captures not the agreement but the degree to which scores on repeated trials relate in a linear manner.[18] An ICC value of >0.80 represents high reliability.[19]

The MDC was calculated based on standard error of measurement (SEM) according to the following formulae[20]:

$$MDC = 1.96 \times \sqrt{2} \times SEM$$

$$SEM = SD_{\text{all testing scores}} \times \sqrt{(1 - r)}$$

The $\sqrt{2}$ is used to account for the underlying extra uncertainty during measurement in two time points. The value 1.96 is the *z* score associated with the 95% confidence level, and *r* is the coefficient of the test–retest reliability, which was estimated by ICC. MDC% was calculated as (MDC/mean) × 100%, presenting the relative amount of random measurement error; the mean is the mean score of all trials. An MDC% of <30% was considered as acceptable and <10% as excellent.[21]

A paired *t* test was conducted for each subtest to examine whether systematic biases existed. The level of significance was set at $p < 0.05$.

Bland–Altman plots were used to visualize the difference and mean score of each pair of measurements.[22] Assuming that the differences follow normal distribution, the limits of agreement (LOAs) lie within $d \pm 1.96 \times SD$, where *d* represents the mean difference between test and retest scores, and SD is the standard deviation of differences of each pairs.

Analyses were carried out using the SPSS 13.0 software (SPSS Inc., Chicago, IL, USA).

Results

A total of 154 patients met the recruitment criteria and completed the first session. However, seven patients did not complete the retest session and were excluded from the data analysis. A total of 132 participants were right-hand dominant and 15 were left-hand dominant. The mean age of the 147 participants was approximately 41 years (SD = 10.1), and 67% of the participants were male. All participants were receiving maintenance medication, taking antipsychotic medication alone (the first three medications were haloperidol, clozapine, and sulpiride), and there were no significant changes in medication in the 1-week study period; the mean dosages of medications are shown in Table 1. In addition, 97 (66.0%) of the participants were receiving regular occupational therapy. The other participants were involved in sheltered employment, vocational training, supported employment, and competitive employment. Further demographic information of the participants is given in Table 1.

Table 1 Characteristics of the clients with schizophrenia (n = 147).

Characteristic	
Gender (male/female)	98/49
Age [mean year (SD)]	41.2 (10.1)
Onset duration [mean year (SD)]	25.0 (6.7)
Schizophrenia subtypes	
295.10 Disorganized type	18
295.20 Catatonic type	1
295.30 Paranoid type	106
295.60 Residual type	16
295.90 Undifferentiated type	6
Education	
Elementary school	14
Junior high school	43
Senior high school	70
University and above	20
Marital status	
Single	108
Married	24
Divorced	15
Vocational status	
No	97
Sheltered employment	17
Vocational training	29
Supported employment	3
Competitive employment	1
Antipsychotic/anticholinergic medication (CPZ) [percentage (mean dosage)]	
Haloperidol	33.3 (644.9)
Clozapine	21.1 (619.4)
Sulpiride	6.8 (380.0)

CPZ = chlorpromazine equivalent doses (mg/d); SD = standard deviation.

The reliability indices of the Purdue pegboard test are listed in Table 2. The ICCs for the five subtests of the Purdue pegboard test between successive sessions were between 0.73 and 0.88, indicating moderate-to-good test–retest reliability. The significance of paired *t* tests of the ND hand, D + ND + B hands, and assembly subtests was significant.

The MDC (MDC%) was 3.0 (22.9%) for the D hand subtest, 3.1 (26.1%) for the ND hand subtest, 3.0 (31.7%) for the B hands subtest, 6.1 (17.7%) for the D + ND + B hands subtest, and 8.5 (35.3%) for the assembly subtest. All MDC% of the five subtests were between 35.3% and 17.7%, representing limited-to-good random measurement error.[19]

In Figs. 1 and 2, the differences in scores between two successive sessions of the D hand and ND hand subtests are respectively plotted against the mean scores. The LOAs ranged from 3.18 to −2.75 for the D hand subtest, and from 3.49 to −2.5 for the ND hand subtest.

Discussion

Hand dexterity is closely related to one's occupation. This study contributes to the applicability and interpretability of a hand dexterity assessment instrument, the Purdue pegboard test, in several aspects. We found that the ICCs of

Table 2 Reliability indices of the Purdue pegboard test (n = 147).

Subtest	First test	Second test	Difference	p of paired t	ICC (95% CI)	SEM	MDC(%)
	Mean (SD)	Mean (SD)	Mean (SD)				
D hand	12.9 (2.3)	13.1 (2.5)	0.2 (1.5)	0.086	0.80 (0.73–0.85)	1.1	3.0 (22.9)
ND hand	11.7 (2.5)	12.2 (2.4)	0.5 (1.5)	<0.001*	0.79 (0.70–0.85)	1.1	3.1 (26.1)
B hands	9.4 (2.1)	9.5 (2.3)	0.1 (1.6)	0.536	0.73 (0.65–0.80)	1.1	3.0 (31.7)
D + ND + B	34.1 (6.3)	34.8 (6.4)	0.7 (3.0)	0.002*	0.88 (0.84–0.92)	2.2	6.1 (17.7)
Assembly	23.8 (7.7)	24.6 (7.9)	0.8 (4.3)	0.018*	0.84 (0.79–0.89)	3.1	8.5 (35.3)

B hands = both hands; CI = confidence interval; D hand = dominant hand; ICC = intraclass correlation coefficient; MDC = minimal detectable change; ND hand = non-dominant hand; SD = standard deviation; SEM = standard error or measurement.
*Significant differences ($p < 0.05$).

the five subtests of the Purdue pegboard test were between 0.73 and 0.88, indicating moderate-to-good test–retest reliability on patients with schizophrenia.[10,19] These results are consistent with the results from healthy aging individuals,[10] and slightly lower than those from students and persons with multiple sclerosis.[11,19] In addition, in the Bland–Altman plots, no systematic trends were found; i.e., the means were not associated with the differences of each pair of measurements.[22] In general, these results suggest that the Purdue pegboard test is a reliable measure for tracking the change in hand dexterity of patients with schizophrenia over time.

For clinical use, our study determines the magnitude of the random measurement error of the Purdue pegboard test when administrated to patients with schizophrenia. We found that the MDCs of the five subtests of the Purdue pegboard test were 3.0, 3.1, 3.0, 6.1, and 8.5 for the D hand, ND hand, B hands, D + ND + B hands, and assembly subtests, respectively. This finding indicates that only a change between repeated measurements greater than the MDC (e.g., 3.0 points for the D hand subtest) could be viewed, with 95% certainty, as a real change.[23] Thus, our findings can help clinicians interpret the change scores more reasonably.

However, the MDC% of all the subtests, except that of the D + ND +B subtest, were higher than 20% (and even higher than 30% for the B hands and assembly subtests), representing substantial random measurement error.[21] These results indicate that to be regarded as true change (or beyond measurement error), a change in Purdue pegboard test should be relatively large. In addition, hand dexterity seems to be more trait-like and less responsive to neuroleptic treatment.[24] These observations may result in few patients being able to achieve greater change scores of the Purdue pegboard test than the MDC. Thus, the clinical utility of the Purdue pegboard test is threatened. To improve the clinical utility of the Purdue pegboard test, we can try to decrease the value of MDC (by increasing the trial number or rater training).[23] It is necessary for clinicians to select a measure with a small MDC to detect real change in the hand dexterity of patients with schizophrenia disorders.

From a statistical perspective, the MDC can also be regarded as the threshold to identify a statistically significant change in an individual patient. Namely, a change score in an

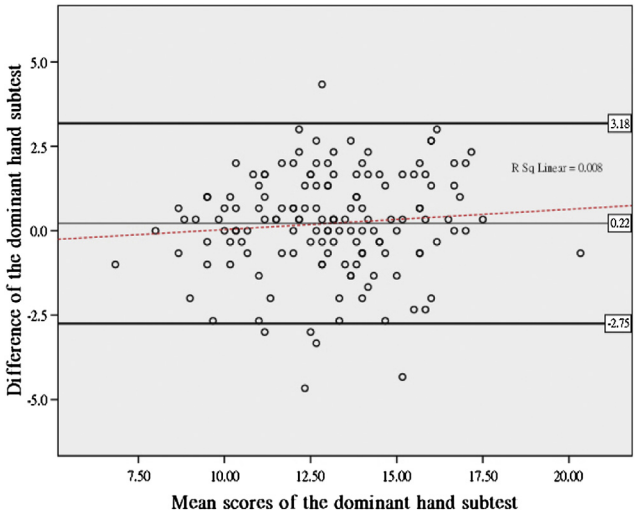

Figure 1 Bland–Altman method for plotting the difference in scores against the mean scores of the dominant hand subtest of the Purdue pegboard test. The two bold lines define the limits of agreement (mean of difference ± 1.96 SD). SD = standard deviation.

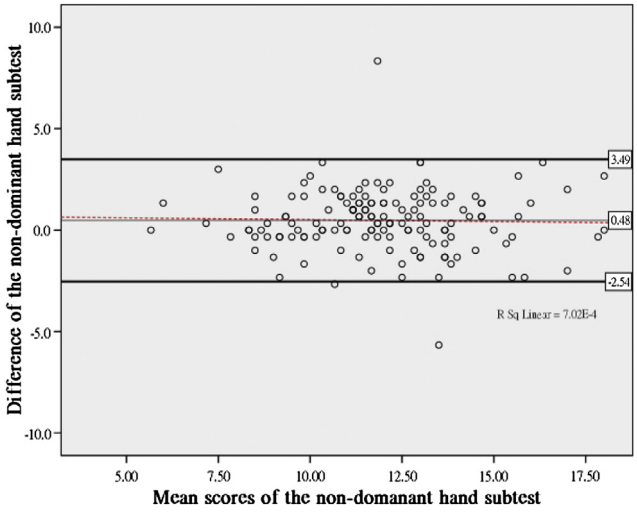

Figure 2 Bland–Altman method for plotting the difference in scores against the mean scores of the nondominant hand subtest of the Purdue pegboard test. The two bold lines define the limits of agreement (mean of difference ± 1.96 SD). SD = standard deviation.

individual patient between two successive measurements exceeding the MDC can be viewed as a change with statistical significance.[25] Therefore, clinicians can use the MDC to determine whether an individual patient has made significant improvement in a clinical setting.

For group comparison (research purposes), the individual-level MDC ($MDC_{individual}$) can be modified to group-level MDC (MDC_{group}), depending on the sample size (n), using following equation: $MDC_{group} = MDC_{individual}/\sqrt{n}$.[26] However, the MDC_{group} is usually not a concern, owing to the substantial sample size in research settings. For instance, if the $MDC_{individual}$ of the D hand subtest is 3.0, the MDC_{group} will be just 0.6 (for $n = 30$), which is too trivial to be of concern.

Researchers can also use the $MDC_{individual}$ as the threshold to report the proportion of a study sample that has achieved real change (improvement or deterioration).[20] Researchers usually report the mean change of the whole study group. The results of significant mean changes within a study group, however, do not guarantee that all patients in the group achieve significant improvement. Thus, reporting the proportion of the study group that achieves or exceeds the $MDC_{individual}$ helps researchers translate their research results to the clinical context and thus enhances the applicability of research outcomes.

Regarding the test procedure, we administered the Purdue pegboard test to patients in groups of two to five persons. There might be a difference between testing the participants in groups or individually. To our knowledge, no studies to date show any differences between testing the participants in groups or individually. However, we cannot exclude the possible effect of group dynamics on the hand function testing of the participants. The difference between testing the participants in groups or individually might need to be examined.

In addition, prospective users could recalculate the value of MDC according to the confidence level that they select. In this study, we used the 95% confidence level (z score = 1.96); however, users could choose another confidence level (e.g., 90%, z score = 1.645) in response to their needs.

Two limitations in the research should be addressed. First, our sample was a convenience sample, with an uneven proportion in gender (male/female = 2/1), as well as in chronic and relatively stable stages. These characteristics of the sample may threaten the generalizability of our findings. Future studies could recruit more patients with more equal gender and illness status distributions to further validate our results.

Second, systematic biases, which may result from the practice effect, were noted in the ND, D + ND + B, and assembly subtests, a fact that may affect the accuracy of our findings. However, the ratios of mean difference to mean scores of test–retest measurements of the five subtests were about 1.1–4.2%, which seems to be too small to be of concern.

In brief, our results reveal that the Purdue pegboard test has moderate-to-good test–retest reliability but substantial random measurement error. These findings should enable clinicians and researchers to monitor and interpret the changes in the hand dexterity of patients with schizophrenia more accurately and confidently.

Acknowledgments

This study was funded in part by Kai-Suan Psychiatric Hospital research project KSPH-2009-03 in Taiwan. The Kai-Suan Psychiatric Hospital had no further role in study design, collection, analysis and interpretation of data, composition of the report, or the decision to submit the paper for publication. We also thank the patients who generously agreed to participate in this clinical study.

References

1. Rahman H. Journey of providing care in hospice: perspectives of occupational therapists. *Qual Health Res* 2000;**10**:806–18.
2. Wolff AL, O'Driscoll GA. Motor deficits and schizophrenia: the evidence from neuroleptic-naive patients and populations at risk. *J Psychiatry Neurosci* 1999;**24**:304–14.
3. Falk-Kessler J, Bear-Lehman J. Hand function in persons with chronic mental illness: a practice concern. *Occup Ther Ment Health* 2003;**19**:61–7.
4. Lehoux C, Everett J, Laplante L, Emond C, Trepanier J, Brassard A, et al. Fine motor dexterity is correlated to social functioning in schizophrenia. *Schizophr Res* 2003;**62**:269–73.
5. Michon HW, Kroon H, van Weeghel J, Schene AH. The Generic Work Behavior Questionnaire (GWBQ): assessment of core dimensions of generic work behavior of people with severe mental illnesses in vocational rehabilitation. *Psychiatr Rehabil J* 2004;**28**:40–7.
6. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;**86**:420–8.
7. Kovacs FM, Abraira V, Royuela A, Corcoll J, Alegre L, Tomas M, et al. Minimum detectable and minimal clinically important changes for pain in patients with nonspecific neck pain. *BMC Musculoskelet Disord* 2008;**9**:43–51.
8. Buddenberg LA, Davis C. Test–retest reliability of the Purdue pegboard test. *Am J Occup Ther* 2000;**54**:555–8.
9. Desrosiers J, Hebert R, Bravo G, Dutil E. The Purdue pegboard test: normative data for people aged 60 and over. *Disabil Rehabil* 1995;**17**:217–24.
10. Gallus J, Mathiowetz V. Test–retest reliability of the Purdue pegboard for persons with multiple sclerosis. *Am J Occup Ther* 2003;**57**:108–11.
11. Guarnaccia VJ, Daniels LK, Sefick WJ. Comparison of automated and standard administration of the Purdue pegboard with mentally retarded adults. *Percept Mot Skills* 1975;**40**:371–4.
12. Jones E, Hanly JG, Mooney R, Rand LL, Spurway PM, Eastwood BJ, et al. Strength and function in the normal and rheumatoid hand. *J Rheumatol* 1991;**18**:1313–8.
13. Gliner JA, Morgan GA, Harmon RJ. Measurement reliability. *J Am Acad Child Adolesc Psychiatry* 2001;**40**:486–8.
14. Hobart J. Rating scales for neurologists. *J Neurol Neurosurg Psychiatry* 2003;**74**:iv22–v26.
15. Tiffin J, Asher EJ. The Purdue pegboard; norms and studies of reliability and validity. *J App Psychol* 1948;**32**:234–47.
16. Yancosek KE, Howell D. A narrative review of dexterity assessments. *J Hand Ther* 2009;**22**:258–69. quiz 270.
17. Tiffin J. Lafayette Instrument Company. *Purdue pegboard quick reference guide: test administrator's manual* 1999. Lafayette, IN.
18. Liu J, Drutz C, Kumar R, McVicar L, Weinberger R, Brooks D, et al. Use of the six-minute walk test poststroke: is there a practice effect? *Arch Phys Med Rehabil* 2008;**89**:1686–92.
19. Baumgartner TA, Chung H. Confidence limits for intraclass reliability coefficients. *Meas Phys Educ Exerc Sci* 2001;**5**:179–88.
20. Haley SM, Fragala-Pinkham MA. Interpreting change scores of tests and measures used in physical therapy. *Phys Ther* 2006;**86**:735–43.

21. Smidt N, van der Windt DA, Assendelft WJ, Mourits AJ, Devill WL, de Winter AF, et al. Interobserver reproducibility of the assessment of severity of complaints, grip strength, and pressure pain threshold in patients with lateral epicondylitis. *Arch Phys Med Rehabil* 2002;**83**: 1145—50.
22. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;**1**:307—10.
23. Schreuders TAR, Roebroeck ME, Goumans J, van Nieuwenhuijzen JF, Stijen TH, Stam HJ. Measurement error in grip and pinch force measurements in patients with hand injuries. *Phys Ther* 2003;**83**:806—15.
24. Strauss ME, Buchanan RW, Hale J. Relations between attentional deficits and clinical symptoms in schizophrenic outpatients. *Psychiatry Res* 1993;**47**:205—13.
25. Jette AM, Tao W, Norweg A, Haley S. Interpreting rehabilitation outcome measurements. *J Rehabil Med* 2007;**39**:585—90.
26. de Vet HC, Bouter LM, Bezemer PD, Beursken AJ. Reproducibility and responsiveness of evaluative outcome measures. Theoretical considerations illustrated by an empirical example. *Int J Technol Assess Health Care* 2001;**17**:479—87.